ABSTRACT OF THE DISCLOSURE

Provided are methods, apparatus and computer programs for improved data storage and management. The invention can be implemented in a replacement or add-on to existing operating system file systems. Files in a file system are separated into a set of information components and then all information components of the file system are analyzed to identify duplication of information content. When information components with duplicate content are identified, duplicates are deleted from physical storage and indexes are generated to reflect inclusion of the retained copy of an information component in a plurality of different files. Improvements to content searching is enabled, since relevant components can be identified without retrieving whole files and since search results will include fewer duplicate results.